PPPA 6002: Research Methods & Applied Statistics

Fall 2018
Wednesday
• 12:45 in 1776 G, Rm C-110 (Section 11)
• 6:10 in Duques Hall, Rm 259 (Section 12)
Thursday
• 12:45 in Tompkins Hall, Rm 206 (Section 10)
• 6:10 in Science & Engineering, Rm 7040 (Section 13)

Instructors
Bill Adams (adams@gwu.edu)
Office: MPA Bldg. 601-D Phone: 202-994-7494
Part 1 hours: 4:15-5:45 Wednesdays and Thursdays
Chris Carrigan (ccarrigan@gwu.edu)
Office: MPA Bldg. 601-K Phone: 202-994-5583
Part 2 hours: 4:15-5:45 Tuesdays and Wednesdays

Lab (Rome B104)
Fran Alba (falba@gwmial.gwu.edu)
Sarah Crump (crumps@gwmial.gwu.edu)

Textbooks & Software
• SPSS (available in GW computer labs).
• Blackboard: weekly readings, videos, assignments.

Assignments and Grades
Part 1: Research methods with Bill Adams:
• Mini-quiz for Part 1 counts 10%.
• Exam for Part 1 counts 25%.
• Additional assignments in Part 1 count 15%.
Part 2: Applied Statistics with Chris Carrigan:
• Policy research project in Part 2 counts 20%.
• Exam for Part 2 counts 25%.
• Additional assignments in Part 2 count 5%.

• Class attendance is crucial and is required; a penalty may apply for more than one unexcused absence.
• If for any reason a class is missed, in whole or in part, please obtain all announcements and assignments (especially those due by the next session), along with class notes and handouts, from your class colleagues and the course Blackboard site before the next class.

Learning Objectives
PPPA 6002 focuses on practical skills for conducting and evaluating empirical and quantitative research, plus provides a brief look at qualitative methods. The course explores the strengths and weaknesses of experimental (RCT), quasi-experimental, and nonexperimental research. It also covers the widely used statistical software, SPSS, and foundational univariate, bivariate, and multivariate statistics.

In particular, the course targets as learning objectives key research skills including how to:

Skill 1: Conduct and evaluate survey research
Skill 2: Conduct and evaluate RCTs
Skill 3: Conduct and assess other methods such as NEC, time series designs, and correlational designs
Skill 4: Conduct and evaluate qualitative studies
Skill 5: Conduct statistical analyses using SPSS
Skill 6: Analyze basic univariate statistics
Skill 7: Implement bivariate statistical techniques including t-tests, chi square, correlation, and regression
Skill 8: Implement multivariate statistical techniques such as multiple regression and partial tables analysis
Skill 9: Prepare a policy research study summarizing statistical analyses for a non-technical audience
Part 1: Additional Information (Adams)

**Blackboard:** All weekly Part 1 course information including readings, videos, PowerPoints, and assignments will be posted on the course Blackboard site.

**Office hours:** You can always stop by during office hours, but to make sure you do not have to wait, please sign up at bill-adams.youcanbook.me by noon that day.

**Weekly Steps:**

1. To stay on top of the material, each week before class it will be valuable to internalize concepts listed on the syllabus and covered in the prior session. Identify any that need clarification.

2. Study assigned chapters, links, and other supplemental readings that were introduced in the prior class session. If you also want to skim over readings for the new material, that is entirely optional.

3. Be sure to go over each week's readings and videos in Blackboard.

4. Submit weekly worksheet answers via Blackboard (at least 1 hour before class). Answers should be concise but usually more than a few words.

5. Be sure to bring a copy of the worksheet questions as well as your answers to class for discussion.

**Grading:** Lowest overall Part 1 grade (no rounding) for an A is 94.00; A- 90.00; B+ 87.00; B 83.00; B- 80.00; C+ 77.00; C 73.00; and C- 70.00.

**Part 1 Examination:** This exam is closed-book (no extra notes and no calculator). It consists of short objective questions focusing on understanding and applying concepts in the syllabus.

**Lab Sessions:** During Part 1, the lab period will be used as shown on the next page. Other weeks you are urged to use the lab time to do the class readings and review videos, complete your weekly worksheets, and/or meet with your study group.

Part 2: Additional Information (Carrigan)

**Blackboard:** All Part 2 course information including class notes, additional readings, and homework assignments will be posted on the course Blackboard site.

**Office Hours:** To sign up for office hours, please visit christopercarrigan.youcanbook.me. Signing up for an available time slot will ensure that you will not need to share that time with another student.

**Textbook:** The textbook for Part 2 (Healey's *Statistics, 10th edition*) is optional. The chapters associated with the material we cover each week are listed in the course detail section below.

**SPSS:** SPSS is the statistical software used in the course. In addition to Rome B104, where lab sessions will take place, SPSS is available in the computer labs at Gelman Library and the Hall of Government. While it is widely used, some more advanced statistics courses at GW (including PPPA 6013) utilize other software to allow you to gain experience with a variety of statistical packages. You should not feel that you need to purchase SPSS, but if you do decide you want a copy, it is available at the bookstore.

**Lab Sessions:** During scheduled labs, the class TAs, Fran Alba and Sarah Crump, will hold regular office hours and provide SPSS assistance to help you with your assignments and policy research project. The first lab to introduce SPSS is during week 5. Additional SPSS labs will be held weekly from week 9 through week 13. Attendance is encouraged but not mandatory.

**Homework Assignments:** Part 2 homework assignments will be graded on a check-plus or check-minus system, based on whether the assignment was fully completed. Feel free to work with classmates, but if you do decide to work with other students, please still turn in your own solutions. Responses should be submitted via Blackboard prior to the start of class.

**Policy Research Project:** This project offers you an opportunity to further develop your skills in analyzing data by generating relevant statistics and interpreting them using a dataset of your choice. Papers should be submitted via Blackboard on or before the due date.

**Part 2 Examination:** The exam will not be cumulative. Rather, it will draw exclusively from material covered in Part 2. Details regarding the exam format will be provided closer to the date.
<table>
<thead>
<tr>
<th>Session</th>
<th>6002 Session Topics</th>
<th>Lab</th>
<th>Homework Due</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1: Aug 29-30</td>
<td>Field trends; Research questions; Literature review; Research ethics</td>
<td>TBD</td>
<td>None</td>
</tr>
<tr>
<td>Week 2: Sept 5-6</td>
<td>Measurement validation; Data gathering options</td>
<td>✓ Paper Q&amp;A</td>
<td>✓ #1</td>
</tr>
<tr>
<td>Week 3: Sept 12-13</td>
<td>Question and questionnaire design; Survey sampling systems</td>
<td>✓ Paper Q&amp;A</td>
<td>✓ #2</td>
</tr>
<tr>
<td>Week 4: Sept 19-20</td>
<td>Causal inference and RCT designs</td>
<td>TBD</td>
<td>✓ #3</td>
</tr>
<tr>
<td>Week 5: Sept 26-27</td>
<td>NEC group designs; Time series; Correlational designs</td>
<td>✓ SPSS lab intro</td>
<td>✓ #4</td>
</tr>
<tr>
<td>Week 6: Oct 3-4</td>
<td>Qualitative research; Focus groups Content analysis; Meta-analysis</td>
<td>✓ Quiz</td>
<td>✓ #5</td>
</tr>
<tr>
<td>Week 7: Oct 10-11</td>
<td>Univariate descriptive statistics</td>
<td>✓ Exam review</td>
<td>✓ #6</td>
</tr>
<tr>
<td>Week 8: Oct 17-18</td>
<td><em>Examination (Conclusion of Part 1)</em></td>
<td>Part 2: Sampling distribution (classrooms TBD)</td>
<td></td>
</tr>
<tr>
<td>Week 9: Oct 24-25</td>
<td>Estimation and confidence intervals</td>
<td>✓</td>
<td>None</td>
</tr>
<tr>
<td>Week 10: Oct 31-Nov 1</td>
<td>Hypothesis testing; Difference of means</td>
<td>✓</td>
<td>✓ #7</td>
</tr>
<tr>
<td>Week 11: Nov 7-8</td>
<td>Bivariate regression and correlation</td>
<td>✓</td>
<td>✓ #8</td>
</tr>
<tr>
<td>Week 12: Nov 14-15</td>
<td>Multiple regression</td>
<td>✓</td>
<td>✓ #9</td>
</tr>
<tr>
<td>Week 13: Nov 28-29</td>
<td>Contingency tables; Chi square</td>
<td>✓</td>
<td>✓ #10</td>
</tr>
<tr>
<td>Week 14: Dec 5-6</td>
<td>Partial tables analysis</td>
<td>✓ Exam review</td>
<td>✓ Policy research study due</td>
</tr>
<tr>
<td>Week 15: Dec 12-13</td>
<td><em>Examination (Conclusion of Part 2)</em></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Introduction to 6002, plus:
Field trends; Research Questions;
Literature Review; Research Ethics

Readings: • Patten, Appendix A and C
• Patten, topics 3, 10-19, 24, 83 (Note “topics” not pages)
• Blackboard readings and videos

Broad 20th century research trends
“Traditionalism”
Behavioralism / basic research
Classic model of scientific research steps

Theory; Hypothesis
Independent variable (X); Dependent variable (Y)
Operationalize concepts; operational definition

Applied research / policy analysis
Model of applied research steps

Basic structure of the written research report:
Intro & problem statement; lit review;
methodology; findings; discussion

Writing a problem statement
Main goals of literature review
Some tips for upgrading the literature review

Belmont Report and principles of research ethics
Institutional Review Board (IRB)
Informed consent
Issues with coercion, confidentiality, risks, deception,
vulnerable populations, common courtesy

Data Gathering; Levels of Measurement;
Measurement Validity and Reliability

Readings: • Patten, topics 20, 23, 35-37, 39 (Extra 40-41)
• Blackboard readings and videos

Seeking representative, valid, reliable data
Secondary analysis of data vs. original data collection
Investigate before using existing datasets
Pros/cons of main data gathering options

Levels of measurement:
Nominal, Ordinal, Interval, Ratio
Interval and ratio called scale
Nominal and ordinal called categorical
Important because appropriate statistics vary
depending on the level of measurement

Measurement reliability & measurement validity
Subjective validity: face validity
Criterion validity: concurrent & predictive validity
Unobtrusive measures; multiple measures

Survey Research

Readings: • Patten: topics 25-27, 30-32 (Extra 33)
• Blackboard readings and videos

Principles for designing good individual questions
Ways to filter or minimize "random responses"
Best practices for overall questionnaire flow:
Short intro; easy start; broader to more detailed;
sensitive questions later; demographics at end
Closed-ended vs. open-ended questions
Likert item (strongly agree/agree/disagree/strongly disagree)

Census vs. sample
Random vs. nonrandom samples
Nonrandom (nonprobability) sampling such as
convenience, snowball, and purposive sampling
Simple random sampling
Systematic random sampling
Stratified random sampling
(Proportionate vs. nonproportionate)
Probability-proportional-to-size (PPS) sampling
Sampling frame (source/list used to draw sample)

Nonresponse bias vs. response bias
Evaluating response rates; how high is high enough?

Survey participation as a quick cost-benefit decision
Best practices for improving survey response rates
Weighing the sample results to reflect population
Factors for the optimum size of a completed sample
Benchmark confidence intervals (95% level) for
n=100 (±10%); 600 (±4%); 1100 (±3%)
Causal Inference and RCT Designs

Readings:
• Patten, topics 52-55, 57
• Blackboard readings and video

Three elements of causal inference...
1) X & Y covary; association; concomitant variation
2) X before Y; direction; time sequence; temporal order
3) Rule out Zs; no plausible alternative; nonspuriousness

Correlation does not prove causation!
Post hoc, ergo propter hoc fallacy

Campbell & Stanley’s design diagraming system
Single group posttest only
Single group pretest-posttest (aka before-and-after)
Static group design (nonequivalent comparison design)

Antecedent variables
Intervening variables
Threats to internal validity (partial list):
- History
- Maturation
- Practice
- Instrumentation
- Regression to the mean
- Selection
- Intra-group history
- Attrition/Mortality (and how to deal with attrition)

Randomized, controlled trial (RCT) = true experiment

Elements of an RCT:
(1) random assignment of subjects from pool to groups and (2) random assignment of X to groups

Reason for the power of RCTs:
Comparability of the groups (i.e., only real difference between the groups is X, so X is the best explanation for differences in the groups)

Classic experimental design
(aka pretest-posttest control-group design)

Posttest only experiment
(aka posttest-only control-group design)

RCT variations:
“Control group” may get something
May have more than one X (factorial designs)
Can assign collectivities (instead of individuals)
Groups not always assigned 1:1 (e.g., may be 2:1)

“Intent to treat” analysis includes all those assigned to treatment group regardless of participation

Factorial designs (simple or complex)
Dosage/sensitivity designs
Complex X
Multiple Ys

External validity (generalizability)
Random selection from the relevant population strengthens external validity.
Random assignment from pool of subjects to groups strengthens internal validity.

Reactivity
Hawthorne effects
Placebo
Try to avoid between-group reactivity as well as other types of reactivity (e.g., with X and staff)

Big four sets of validity issues:
- Measurement validity
- Internal validity
- Statistical conclusion validity
- External validity

RCT’s two essential elements

Pool of Subjects
Random Assignment of Subjects to Groups
Random Assignment of X
Treatment Group
Control Group
NEC, Time Series, Correlational Designs

Readings:  • Patten, topics 4-5, 56
  • Blackboard readings and videos

Practical reasons why RCTs may not be conducted
Quasi-experiments (vague term)
Causal-comparative (another term for studies that try to
infer causality when groups not randomly assigned)
Nonequivalent comparison group (NEC) designs
Pretest-posttest nonequivalent comparison design
Posttest only nonequivalent comparison group design

Key internal validity threat to NEC designs: selection
Retrospective matching design (ex post facto
with nonrandom posttreatment matching)
Natural experiments (strict vs. broad usage of term)

Time series (aka longitudinal) research
Why superior to “single group pretest-posttest”?  
Key internal threat to time series study: history
Simple interrupted time series
Reiterative time series; Multiple time series

Deceptive time series charts (truncated base)
Panel data (aka "panel-back") vs. cross-sectional data
Aggregate data (units of analysis are groups)
Retrospective pretests; proxy pretests
Fallacy of time series inferences from a single survey
Ecological fallacy

Process and logic of correlational designs
Key internal threat to correlational studies: selection
  Difficulty in statistically controlling for all Zs
  especially selection threats of motivation and
  self-selection (i.e., specification error, aka omitted
  variable bias); so different control variables can
  produce widely varying results

Overall assessment of causal designs:
Lab experiments: often strong on internal validity
  but weak on external validity
Nonexperimental field studies: often strong on
  external validity but weak internal validity
Field experiments: strong in both internal and
  external validity but often not feasible to conduct

The logic of inferring causality by coupling lab
experiments with nonexperimental field studies

Qualitative Research

• Patten, topics 7-8, 29, 33, 46-51
  • Blackboard readings, including
    “Semi-Structured Interviews”

Qualitative Research
  More exploratory than hypothesis testing
  Small, purposive sample, not large random
  Extended, intense observations or interviews
  Unstructured or semi-structured data gathering
  Essay reports with little or no quantitative data
  Often explore the researchers’ subjective impact

Focus groups purposes:
  Probing attitudes, reaction testing, brainstorming
Focus group steps: recruit relevant people; 10-12;
  1½-2 hours; semi-structured format with mostly
  open-ended topics; neutral facilitator.

Mixed Methods Research
  Using both qualitative & quantitative approaches,
  For example, qual, then quant, then qual.

Content analysis steps:
  Define scope
  Operationalize variables to code
  Refine and test coding system
  Inter-coder reliability testing
  Code content and analyze data

Meta-analysis purpose and strengths
  Steps in conducting a meta-analysis

Focus groups purposes:
  Probing attitudes, reaction testing, brainstorming
Focus group: participant recruitment; focus group
  size; session length and agenda; moderator style;
  and ideal focus group facilities
7 Univariate Descriptive Statistics

Readings:
- Patten, topics 60-64
- Blackboard readings and videos

Good data analysis requires good data, plus awareness that: all summary statistics are reductionist, context dictates interpretation, small differences should not be exaggerated, correlation does not prove causation, start with univariate analysis before multivariate.

Nominal univariate statistics – percent and mode
Interpretation pitfalls include:
- Misleading pictograms; confusing absolute and relative percent; misinterpreting mode as midpoint; and misleading modal composites
Plurality vs. majority
Major measures of central tendency:
- mean and median, plus trimmed mean
Mode (not necessarily a central tendency)

Major measures of dispersion:
- standard deviation and interquartile range
Positive skew (high values pull mean above median)
Negative skew (low values pull mean below median)

Normal curve
± 1 standard deviation = 68.3% of normal curve
± 2 standard deviations = 95.4% of normal curve
± 3 standard deviations = 99.7% of normal curve

Value of examining frequency distribution charts
Descriptive vs. inferential statistics

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>( \mu ) “mu”</td>
<td>( \sigma ) “sigma”</td>
</tr>
<tr>
<td>Sample</td>
<td>( \bar{X} ) “x-bar”</td>
<td>s</td>
</tr>
</tbody>
</table>

Boxplots, stem-and-leaf plots
Histograms, bar charts, pie charts

8 During Class: Examination for Part 1
After Exam: Introduction to Part 2

8 Introduction to Part 2: Sampling Distribution

Readings: Healey, chapter 6

Inferential v. descriptive statistics
Sampling distribution (theoretical distribution of a statistic for all possible sample outcomes of given size, \( n \))
Mean of the sampling distribution = \( \mu \)
Standard deviation of the sampling distribution = \( \sigma \)

Properties of the sampling distribution:
1. \( \mu = \mu \)
2. \( \sigma = \sigma/\sqrt{n} \) (standard error)

Central limit theorem (as sample size grows, the sampling distribution approaches normal regardless of the shape of the population distribution)

9 Estimation and Confidence Intervals

Readings: Healey, chapter 7

Two types of estimates, point and interval
Estimator (used to estimate the population parameter by approximating it)
Good estimators are 1) unbiased and 2) efficient

Formula to construct a confidence interval around a sample mean when do not know \( \sigma \):
\[
c.i. = \bar{X} \pm Z \frac{s}{\sqrt{n}}
\]
where \( Z \) is the critical value

Consider alpha (\( \alpha \)), which is the probability that the interval does not contain the population parameter, and the confidence level (1 - \( \alpha \)) to determine the critical value

Adjust the interval through \( n \) and the confidence level

Formula to construct a confidence interval around a sample proportion:
\[
c.i. = p \pm Z \frac{p(1-p)}{n}
\]
where \( p \) is the sample proportion
Hypothesis Testing; Difference of Means

Readings: Healey, chapters 8 and 9

Hypothesis testing steps:
1. State the assumptions
2. State the null hypothesis ($H_0$) and alternative/research hypotheses ($H_a$)
3. Select the critical value
4. Compute the test statistic
5. Compare the test statistic to the critical value.
   Decide whether to reject or fail to reject $H_0$

Null hypothesis is a statement of no difference, specified in terms of populations. The null is the assumption, but it is never “proven.” Failure to detect a relationship, especially in a small sample, does not mean there is no relationship.

$Z$ (critical) = ±1.96 if want to be 95% confident, associated with $\alpha = 0.05$

Formula to compute the test statistic for a two-sample means test when $n_1 \geq 30$ and $n_2 \geq 30$:

$$Z_{obtained} = \frac{\bar{X}_1 - \bar{X}_2}{\sigma} = \frac{\bar{X}_1 - \bar{X}_2}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Formula to compute the test statistic for a two-sample proportions test (where $\pi = \frac{p_1 + p_2}{2}$):

$$Z_{obtained} = \frac{p_1 - p_2}{\text{SE}(p_1 - p_2)} = \frac{p_1 - p_2}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

Student’s $t$ distribution replaces $Z$ distribution where $df$ (degrees of freedom) is $n - 1$ when $n < 30$ since $s$ is no longer a good estimator of $\sigma$. As $n$ increases, $t$ distribution converges to $Z$ distribution.

Tradeoffs in testing:
1. Type I v. type II error. Lowering $\alpha$ reduces type I error (reject true null) but increases type II error (fail to reject false null)
2. Statistical v. substantive significance. Large samples can show statistical significance for trivial relationships
3. One v. two-tailed tests. A one-tailed test increases the likelihood of rejecting the null by lowering $Z$(critical) but only if theory supports it

Bivariate Regression and Correlation

Readings: Healey, chapter 13

Scatterplot (positive, negative, or no relationship)

Formula for bivariate regression: $Y = a + bX$

where $\hat{Y}$ = predicted value for dependent variable ($Y$ on the regression line), $b$ = slope, and $a$ = intercept.

Regression gives the formula for the straight line that comes closest to the conditional means (average $Y$’s for observations with the same $X$ value).

Slope represents the “magnitude.” The amount of change in $Y$ when $X$ increases by 1 unit.

Residuals represent the difference between the actual and predicted values ($\hat{Y} - Y$).

Regression assumes the relationship is linear. Not appropriate for curvilinear patterns unless the specification of the variables is altered.

Formula to compute the test statistic for the hypothesis test to determine whether there is a relationship between $X$ and $Y$ in the population ($H_0$: $\beta = 0$ where $\beta$ = population slope coefficient) is:

$$t_{obtained} = \frac{b - 0}{\text{SE}(b)}$$

where $\text{SE}(b)$ is the standard error of the slope.

$p$-value = the probability of observing a test statistic equal to or further from the center of the distribution than that obtained if the null is true.

Correlation coefficient ($r$) ranges from -1 to 1 and measures the strength of the relationship. No linear relationship when $r = 0$.

Rough (and somewhat arbitrary) cutoffs for the strength: $-0.3 \leq r \leq 0.3$ (weak); $-0.6 \leq r < -0.3$ or $0.3 < r \leq 0.6$ (moderate); $r < -0.6$ or $r > 0.6$ (strong).

Null hypothesis in test of statistical significance of $r$ is $\rho = 0$ where $\rho$ = population correlation coefficient.
**Multiple Regression**

*Readings: Healey, chapter 15*

Formula for multiple regression:
\[ Y = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_k X_k \]

where, e.g., \( b_1 \) = slope of the linear relationship between \( X_1 \) and \( Y \)

Each slope shows the amount of change in \( Y \) when that independent variable increases by 1 unit:
- holding the other independent variables constant
- controlling for the effects of other independent variables

Coefficient of determination \( R^2 \) measures the proportion of the variation in \( Y \) that can be explained by the regression.

Some limitations of \( R^2 \) even when its value is high:
1. Always increases when add variables (except in the rare case where the additional variable has absolutely no effect)
2. Does not indicate which variables are significant
3. Does not mean that important variables have not been omitted

Adjusted \( R^2 \) does not necessarily increase when a variable is added to the regression.

Dummy (indicator) variables take the values 0 and 1. Used when nominal or ordinal independent variables are included in the regression.

To avoid perfect multicollinearity, add one fewer dummies than the number of categories in the nominal/ordinal variable.

Intercept is the predicted value for the omitted category and slope coefficients are interpreted relative to the omitted category.

**Partial Tables Analysis**

*Readings: Healey, chapter 14*

Partial tables analysis (aka crosstabs with controls) examines the relationship between \( X \) and \( Y \) for each category of \( Z \).

Introducing \( Z(s) \) can change the understanding of the relationship between \( X \) and \( Y \) in various ways.

Relationship can be:
1. Direct (\( Z \) does not alter the \( X-Y \) relationship)
2. Spurious (\( Z \) is the cause of the \( X-Y \) relationship)
3. Intervening (\( Z \) is the link between \( X \) and \( Y \))
4. Suppressor (absence of \( Z \) is masking the \( X-Y \) relationship)
5. Interacting (\( X-Y \) relationship changes across categories of \( Z \))

**Contingency Tables; Chi Square**

*Readings: Healey, chapters 11 and 12 (through p. 317)*

When both variables are nominal or ordinal, can create a “contingency” table (aka crosstab) with the independent variable in the columns and compute the “conditional” percentages.

Maximum percentage-point difference in the conditional percentages represents the relationship magnitude.

Rough (and somewhat arbitrary) cutoffs for the relationship strength: 0-10% points (weak); >10-30% points (moderate); >30-100% points (strong)

Chi square distribution with \( v \) degrees of freedom (df) is the sum of \( v \) squared independent standard normal random variables:
\[ \chi^2 = Z_1^2 + Z_2^2 + \ldots \]

Degrees of freedom for \( \chi^2 \) (critical \( (r-1)(c-1) \) where \( r \) is the number of rows and \( c \) is the number of columns in the contingency table.

Formula to compute the chi square test of independence:
\[ \chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} \]

where \( f_o \) = observed cell frequencies and \( f_e \) = cell frequencies that would be expected if the variables are independent.

Formula to determine the expected frequencies:
\[ \frac{f_0}{f_e} = \frac{\text{row marginal} \times \text{column marginal}}{n} \]

**Examination for Part 2**
Standard Policies

1. The Syllabus: This syllabus is a guide to the course. Sound educational practice requires flexibility and the instructor may therefore, at her/his discretion, change content and requirements during the semester.

2. Incompletes: A student must consult with the instructor to obtain a grade of "I" (incomplete) no later than the last day of classes in a semester. At that time, the student and instructor will both sign the CCAS contract for incompletes and submit a copy to the School Director. Please consult the TSPPPA Student Handbook or visit the website for the complete CCAS policy on incompletes.

3. Submission of Written Work Products Outside the Classroom: It is the responsibility of the student to ensure that an instructor receives each written assignment. Students can submit written work electronically with the express permission of the instructor.

4. Policy on Late Work: All work must be turned in by the assigned due date in order to receive full credit for that assignment, unless an exception is expressly made by the instructor.

5. Academic Honesty: The GW Code of Academic Integrity is at studentconduct.gwu.edu/code-academic-integrity. All exams and other graded work products are to conform to the Code. It defines "academic dishonesty" as "cheating of any kind" and "misrepresenting one's own work, taking credit for the work of others without crediting them and without appropriate authorization, and the fabrication of information."

6. Changing Grades after Completion of the Course: No changes can be made in grades after the conclusion of the semester, other than in cases of clerical error.

7. Religious Holidays: Religiously observant students should notify the instructor the first week of classes regarding any session that will be missed; the courtesy of an absence without penalty will be extended.

8. Accommodation for Students with Disabilities: To receive accommodations on the basis of disability, please provide documentation from the GW’s Disability Support Services, Rome Hall 102 (202-994-8250). See also: disabilitysupport.gwu.edu/.

9. Mental Health Services: This GW office offers 24/7 assistance to address students’ personal, social, career, and study skills problems, along with emergency mental health consultations and counseling services as well as referrals. See: counselingcenter.gwu.edu

10. Community Values: Higher education works best when it becomes a vigorous and lively marketplace of ideas in which all points of view are heard. Free expression in the classroom is an integral part of this process. Higher education also works best when we approach the enterprise with empathy and respect for others, irrespective of their ideology, political views, or identity. We value civility because that is the kind of community we want, and civility enables more effective intellectual exploration and growth.

---

GW Bulletin Course Description (bulletin.gwu.edu/courses/pppa)

PPPA 6002. Research Methods and Applied Statistics

Development of skills and knowledge for conducting original research and critically evaluating empirical studies. Various research designs and data collection techniques are examined. Focus on computerizing data sets for quantitative analysis, analyzing strength of relationships, selecting appropriate statistical techniques, and testing statistical hypotheses.

Average Minimum Independent Weekly Work:

In addition to the average of three hours weekly of direct instruction in class and the computer lab, this course requires a minimum weekly average of 5-6 hours of independent reading, research, and learning.